ELSEVIER

Social Science & Medicine 64 (2007) 1392–1406

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

Linked versus unlinked estimates of mortality and length of life by education and marital status: Evidence from the first record linkage study in Lithuania

Vladimir M. Shkolnikov[a,*], Domantas Jasilionis[a], Evgeny M. Andreev[a], Dmitri A. Jdanov[a], Vladislava Stankuniene[b], Dalia Ambrozaitiene[c]

[a] *Max Planck Institute for Demographic Research Rostock, Germany*
[b] *Demographic Research Centre, Institute for Social Research, Lithuania*
[c] *Statistics Lithuania, Lithuania*

Available online 29 December 2006

Abstract

Earlier studies have found large and increasing with time differences in mortality by education and marital status in post-Soviet countries. Their results are based on independent tabulations of population and deaths counts (unlinked data). The present study provides the first census-linked estimates of group-specific mortality and the first comparison between census-linked and unlinked mortality estimates for a post-Soviet country. The study is based on a data set linking 140,000 deaths occurring in 2001–2004 in Lithuania with the population census of 2001. The same socio-demographic information about the deceased is available from both the census and death records. Cross-tabulations and Poisson regressions are used to compare linked and unlinked data. Linked and unlinked estimates of life expectancies and mortality rate ratios are calculated with standard life table techniques and Poisson regressions. For the two socio-demographic variables under study, the values from the death records partly differ from those from the census records. The deviations are especially significant for education, with 72–73%, 66–67%, and 82–84% matching for higher education, secondary education, and lower education, respectively. For marital status, deviations are less frequent. For education and marital status, unlinked estimates tend to overstate mortality in disadvantaged groups and they understate mortality in advantaged groups. The differences in inter-group life expectancy and the mortality rate ratios thus are significantly overestimated in the unlinked data. Socio-demographic differences in mortality previously observed in Lithuania and possibly other post-Soviet countries are overestimated. The growth in inequalities over the 1990s is real but might be overstated. The results of this study confirm the existence of large and widening health inequalities but call for better data.

Keywords: Socio-economic inequalities; Differential mortality; Health; Lithuania; Baltic countries

Introduction

In the countries of the former USSR, generally high mortality co-existed with large inequalities in mortality in relation to major socio-demographic characteristics, such as education, marital status, and

*Corresponding author. Tel.: +49 3812081147; fax: +49 3812081447.
E-mail addresses: shkolnikov@demogr.mpg.de (V.M. Shkolnikov), jasilionis@demogr.mpg.de (D. Jasilionis), Andreev@demogr.mpg.de (E.M. Andreev), jdanov@demogr.mpg.de (D.A. Jdanov), vladast@ktl.mii.lt (V. Stankuniene), dalia.ambrozaitiene@stat.gov.lt (D. Ambrozaitiene).

0277-9536/$ - see front matter
doi:10.1016/j.socscimed.2006.11.014

ethnicity (Andreev & Dobrovolskaya, 1993; Kalediene & Petrauskiene, 2000, 2005; Leinsalu, Vågerö, & Kunst, 2003, 2004; Shkolnikov, Deev, Kravdal, & Valkonen, 2004; Shkolnikov, Leon, Adamets, Andreev, & Deev, 1999). Since the early 1990s, a widening in inter-group mortality gaps has been observed in countries of the former Soviet Union and Eastern and Central Europe. The difference in life expectancy at age 25 between the highest and lowest education groups grew between 1989 and 2000 from 8.6 to 13.1 years for Estonian males (Leinsalu et al., 2003). A similarly striking widening in mortality inequalities was identified in Lithuania and Russia (Kalediene & Petrauskiene, 2005; Shkolnikov et al., 2004). The increase in mortality differences by education was less significant in Hungary and only very moderate in the Czech Republic (Blazek & Dzúrová, 2000; Klinger, 2001). Mortality decrease in Estonia in the late 1990s was entirely determined by the higher educational group. The mortality of the rest of the population, by contrast, increased (Leinsalu et al., 2003).

These results are of major significance to public health, but one should keep in mind that in all of these studies group-specific mortality estimates are based on so-called "unlinked data". That is to say that mortality rates were calculated from two independent tables of deaths and population by age and group. Such data potentially can be biased (Valkonen, 1993; Vallin, 1979) due to possible differences between self-reported information from the census and death certificate information provided by proxy informants (relatives). The most reliable data on group-specific mortality are provided by studies linking individual deaths and census records (Valkonen, 1993). However, only few cohort studies of this type have been completed in the former USSR; these are small and nationally unrepresentative (Plavinski, Plavinskaya, & Klimov, 2003; Shkolnikov et al., 2004).

The magnitude of the numerator-denominator bias in Western countries has been estimated in several studies. Comparisons of self-reported and proxy-reported education, race, and occupation, and an examination of the reliability of information indicated in death records were performed in the USA and Great Britain. The results are contradictory. Two studies have shown misreporting in death certificates that was relatively insignificant and thus could not have produced a substantial effect on the group-specific mortality estimates (Goldblatt, 1989; Marmot & McDowall, 1986; Rosamond et al., 1997). Other studies, however, have found significant disagreement (up to 30–40%) between the self-reported data in the survey and the follow-up death registry data (Shai & Rosenwaike, 1989; Sorlie & Johnson, 1996).

Up to now, census-linked mortality data has not been available for any Eastern or Central European country except Bulgaria (Kohler, 2001; Kunst et al., 2004) and the validity of information on the socio-demographic status in death records has never been examined. Some authors have made assumptions about the possible direction of the numerator-denominator bias. Andreev and Dobrovolskaya (1993) wrote that an over-estimation of education in the Russian death records was particularly significant, whereas the similar promotion was less pronounced in the census data. They suggested that an understatement of real mortality differences by the unlinked data is more probable than their overstatement. Later analyses of this paper will show that this statement could not hold true.

Lithuanian data on mortality by socio-demographic group in 2001–2004 provide a unique opportunity for the direct comparison of linked and unlinked estimates of group-specific mortality by the simultaneous use of the same socio-demographic characteristics from the census and the death records.

In the present study, we estimate the numerator-denominator bias and provide unbiased data on mortality differentiation by socio-demographic group. We begin from a systematic analysis of disagreements between the census and the death registry information in reporting the education and marital status of the deceased. We address the problem of operating with deaths that could not be linked to the census, and propose a solution that allows the inclusion of these deaths into the calculation of group-specific mortality. Then, we produce group-specific estimates of life expectancy and mortality rate ratios and compare those derived from the census-linked data with those based on the census-unlinked data. Finally, we consider the direction and magnitude of the numerator-denominator bias in life expectancy and mortality rate ratios and examine to what extent it distorts socio-demographic differences in these mortality indicators.

Data and methods

Data

This study uses a data set from Statistics Lithuania. It is based on all records from the 2001

Population and housing census and all death and emigration records for the period between April 6, 2001 and December 31, 2004. Information about external migration was drawn from the Population Register database. It is not fully complete since a part of emigrants have not declared their departure (Stankuniene, Mikulioniene, & Baublyte, 2002).

The census of 2001 was carried out on the 6th of April 2001 and includes all permanent residents on the territory of Lithuania. Lithuanian citizens who had been staying abroad for a maximum of 1 year at the time of the census were also enumerated, whereas those who had been staying abroad for a longer time period were not counted. Person-years at risk were estimated by adding up the years of persons living in Lithuania between July 1, 2001 and December 31, 2004. For individuals who died or emigrated, the exposure time was censored at the date of death or emigration.

Death records of all individuals (excluding those born after the census) who died between July 1, 2001 and December 31, 2004 were used for the analyses. According to the WHO, the quality of these data is high (Mathers, Fat, Inoue, Rao, & Lopez, 2005). However, there could be some problems related to the causes of death. Earlier studies, found some over-reporting of cardiovascular diseases and under-reporting of cancers, external causes, and alcohol-related deaths (Petrauskienė, Bierontas, Kalėdienė, & Zaborskis 1996; Stalioraityte, Pangonyte, & Kazlauskaite, 2005). The proportion of misclassifications increased with age and was more frequent in small towns and rural areas. The same studies showed that corrections of causes of death did not distort the principal cause-of-death pattern in a significant way.

The aforementioned potential deficiencies of cause of death statistics concern medical death certificates issued by medical and forensic institutions. However, death certificates undergo further examination after their submission to Statistics Lithuania. The statistical office performs a validation of the diagnoses and respective ICD-10 codes in the medical death certificates on a monthly basis. All questionable cases are further checked with the respective medical or forensic institutions. Only after that, the cause of death data are included into the final death database.

Quality of the cause-of-death data can potentially vary by population group. However, the corresponding effects are likely to be minor. It should be recognized that we are analysing cause of death patterns in a small country (total population of 3.5 million as of 2001) with a centralized death certification and coding system. The relatively high autopsy rates (about 30%) in Lithuania also contribute towards diminishing the potential inter-group heterogeneity in the cause-of-death quality. In addition, we consider that potential discrepancies should be negligible in terms of a few broad groups of causes of death.

Seven most important groups of causes of death based on the WHO ICD-10 were selected for analysis: circulatory diseases (I00-I99), neoplasms (C00-D48), external causes (but poisoning by alcohol) (V01-X44, X46-Y89), causes directly related to alcohol such as liver cirrhosis, alcoholic psychosis, chronic alcoholism, poisoning by alcohol (F10.5, F10.0–F10.4, F10.6–F10.9, K70, K74, X45), infectious and respiratory system diseases (A00-B99, J00-J99), ill-defined causes of death (R00-R53, R55-R94, R96-R99), and other causes of death.

Linkage procedures

Linkage between the census, death and migration records was accomplished using personal identification numbers (PINs) as unique identifiers for the same individuals. PINs consist of 11 digits which allow to identify individuals in a unique way: the first digit denotes sex; six successive digits refer to the date of birth (year, month, and day), whereas other three numbers signify the serial number of the birth in the sequence of births of the respective day; and the last digit carries a special control code. Introduced in 1992, PINs have been used in the 2001 census and in the current vital records. Using PINs instead of other possible identifiers such as addresses, names, and other individual characteristics allowed to avoid potential discrepancies due to coding errors in both census and death records. The use of PINs makes a false linkage between census records and vital events almost entirely impossible.

Prior to the main study, we conducted a feasibility study checking the validity of PINs as a means for the linkage between the death and census records. 60,267 or 94% of the total number of 64,395 of deaths registered between June 1, 2001 and December 31, 2002 were successfully linked to the census records. In addition, this study showed very high agreement between the PIN-linked census and death records in respect to sex, date of birth, and age (98%). Almost all of the remaining 2% disagreement was related to misprints in the month and the day of birth in death certificates. The disagreements in year of birth and sex were very rare (0.2% and 0.07% of all linked records, respectively).

In the present study, PINs allowed us to match about 95% of the death records to the corresponding census records. Absence of PINs in either or both census or death records (about 0.8% in both cases) and misprints in PINs were the main obstacles for achieving better results. The remaining 5% census-unlinked deaths were incorporated into our database and analysed by applying a special re-distribution procedure (see section on Treatment of unlinked deaths and Appendix A).

The linkage works were performed at Statistics Lithuania by employees who have permission to work with individual-level data. For the purpose of the present study, Statistics Lithuania produced frequency data containing raw death and exposure-time counts by every combination of socio-demographic variables.

Socio-demographic dimensions

With the exception of the census-unlinked death records, information about age and sex was taken from the census records. For the unlinked deaths, we also used age from the death records. We restricted our analyses to ages above 30, assuming that most people have completed their education by this age. In the life table analyses, 5-year age groups (30–34, 35–39, ..., 80–84, 85+) were used.

Education and marital status were available both from the census and the death records. Categories of marital status are identical in the both the census and death records: married, never married, divorced, widowed and unknown.

There are important differences between the census and the death records in coding of education: 11 categories are indicated in the census questionnaire and only five broad categories are given in the death record. In addition, the "census question" about education stresses on the *completed* highest level of education, whereas the "death record question" simply asks about person's education with a list of education categories to be marked. The latter could sometimes result in a usage of incomplete educational categories—e.g. to mark "secondary" education if 9 years of schooling were completed instead of the obligatory 10–12 years.

Taking into account significant transformations in the educational system throughout the 20th century, we applied the following three broad categories of education[1]: higher education (at least 14 years of schooling), secondary education (10–13 years of schooling), lower than secondary or unknown education (up to 9 years of schooling).

The census-based socio-demographic dimensions such as ethnicity (Lithuanian, Russian, Polish, Other, Unknown), and place of residence[2] (urban, rural) were used only as independent variables for predicting misreporting of education and marital status in univariate Poisson regression analyses.

Table 1 presents person-years and deaths by sex, education, and marital status for ages over 30. The data include 3.2 and 4.1 million person-years of population exposure and 72.5 and 65.9 thousand deaths for males and females, respectively. Death and exposure-time numbers are large for every educational category, but they are relatively small for the category of widowed males.

Dealing with missing data

There was no missing information about sex, place of residence and causes of death in our database. However, missing or unknown categories were found for date of birth, education, marital status, and ethnicity.

In the census, missing year of birth (individuals' age) constituted only 335 cases (or 0.01% of records). They were excluded from further analysis. Among the unlinked deaths, missing age was also extremely rare ($n = 11$, 0.08% of the total number of deaths). These deaths were also excluded. Missing month of birth was replaced by the month of July (mid-year) for 810 (0.02%) cases.

As the proportions of the census and death records with an unknown marital status were negligible (0.09% and 0.4%, respectively). For ethnicity these proportions were also very low (0.4% and 0.07%, respectively). The corresponding exposure times and

[1]Higher education refers to university or non-university higher education (confirmed by a certificate of higher education). Secondary education embraces completed general upper secondary or/and vocational/technical school or college (confirmed by a maturity or vocational certificate of upper secondary education). Lower than secondary or unknown education refers to completed general or vocational basic education, and primary school. Additionally, the latter category includes: persons with incomplete primary education (without certificate), literate persons without schooling or illiterate persons, and persons with their education status unknown.

[2]Place of residence refers to the place of actual residence. Two categories—urban and rural area are distinguished.

Table 1
Population exposure and deaths by education and marital status

	Person-years		Deaths[a]	
	Males	Females	Males	Females
Education				
Higher	504,935 (15.7%)	695,415 (17.0%)	5558 (7.7%)	3244 (4.9%)
Secondary	1,809,071 (56.3%)	2,177,307 (53.1%)	24,322 (33.5%)	16,793 (25.5%)
Lower than secondary/unknown	897,693 (28.0%)	1,223,823 (29.9%)	42,612 (58.8%)	45,880 (69.6%)
Marital status				
Married	2,446,060 (76.2%)	2,424,922 (59.2%)	46,089 (63.6%)	17,622 (26.7%)
Never married	322,198 (10.0%)	331,740 (8.1%)	6766 (9.3%)	6341 (9.6%)
Divorced	320,130 (10.0%)	506,231 (12.4%)	8902 (12.3%)	4290 (6.5%)
Widowed	120,525 (3.7%)	830,884 (20.2%)	10,696 (14.7%)	37,606 (57.1%)
Unknown	2786 (0.1%)	2768 (0.1%)	39 (0.1%)	58 (0.1%)
Total	3,211,699 (100%)	4,096,545 (100%)	72,492 (100%)	65,917 (100%)

Lithuania, all ages above 30, July 1, 2001–December 31, 2004.
[a]Census-unmatched deaths were redistributed according to the method described in the data, the methods section and in Appendix A.

deaths were excluded from the mortality calculations. However, the unknown values of all socio-demographic variables were used in the analyses of misreporting patterns (Appendix B).

The proportions of census records with unknown education are 3% and 5% for males and females, respectively. The corresponding categories according to the death records data constituted 0.9% and 1% from the total number of male and female deaths. The census-based unknown group includes the following census categories: "not finished primary", "literate", "illiterate", and "not indicated".

Our decision to merge unknown and lowest educational categories was supported by the fact that more than 80% of the census-based deaths with "unknown" education fell into the lowest educational categories according to the death records. In addition, among these deaths, proportions of external, infectious and ill-defined causes are especially high, which is characteristic of lower socio-economic groups. A similar procedure of adding the deceased and exposure times with unknown education to the lowest educational category has been applied in other countries (see Valkonen, 1993).

Methods of mortality analysis

To estimate the impact of various factors on the disagreement between the census and death record based information on education, and marital status we applied Poisson regression with a misreporting rate as the dependent variable. This analysis was performed separately for each sex and socio-demographic variable. Age was present in all regressions, whereas separate age-adjusted models were estimated to measure the independent effects of causes of death, education, marital status, ethnicity, and the place of residence (urban/rural).

Note that all estimates of life expectancy in this study reflect the level of *period mortality risk*, which differs from the real cohort length of life. This is especially important when it comes to interpreting life expectancies in changeable marital status groups.

Life expectancies at age 30 and temporary life expectancies between the exact ages 30 and 70 were calculated from the census-linked and unlinked data. Exclusion of old ages generally envisages a lower quality of socio-demographic information at these ages. Standard errors and 95% confidence intervals for life expectancies and temporary life expectancies were estimated according to the Chiang (1984) method.

Relative mortality differences for ages 30+ and 30-69 were obtained from the Poisson regressions with mortality as the dependent variable. Separate models (controlling for age) were estimated for relative mortality differences by education and marital status. Results produced by Poisson regressions are expressed in terms of mortality rate ratios and their 95% confidence intervals.

Treatment of unlinked deaths

The proportion of census-linked death records varies from country to country depending on the quality of population registers, completeness in the registration of current demographic events, and the efficiency of linkage instruments. The percentage of linked records close to 100% is seen in the countries of Northern Europe (Huisman et al., 2005; Valkonen, 1993). The first linkage study in Eastern Europe on Bulgaria allowed linking about 93% of deaths between December 5, 1992 and December 31, 1993 to the 1992 census (Kohler, 2001). In our data, 95% of deaths were linked to the census.

The conventional practice in dealing with unlinked deaths has been simply excluding them from the analysis (Doblhammer, Rau, & Kytir, 2005; Kohler, 2001). However, we found that the exclusion of unlinked deaths leads to a considerable elevation of life expectancy at age 30 for the whole population: by 1 year for men and by 0.5 year for women. We also identified clearly non-random patterns in distributions of unlinked deaths by cause of death, education, marital status, and ethnicity (calculations not shown here). Unlinked deaths show significantly higher percentages of deaths due to external and ill-defined causes and more frequently belong to lower socio-demographic groups (such as lower than secondary education or non-married statuses). Therefore, exclusion of the census-unlinked death records would have resulted in distorted cause- and group-specific mortality estimates.

At the same time, including them into calculations of mortality is problematic as the census-based information on socio-demographic status is missing for these deaths. One might think of using information about socio-demographic status from the death records. But this would also be connected with a disadvantage since the resulting mortality estimates would be partly affected by the numerator-denominator bias. However, the availability of duplicated information on education, marital status, and ethnicity from both the census and the death records for the linked deaths allows us to find a feasible solution (see Appendix A for more details).

Results

Differences between the census and the death records in reporting the socio-demographic status

In this section, we examine data on the deceased by comparing death records and census information on education and marital status. The following analyses are restricted to census-linked death records.

Table 2 compares the self-reported educational categories in the census with the same categories in the death records reported by proxy informants. Our data suggest that misreporting of education is very significant and that the patterns are similar for males and females. The percentage of agreement ranges from 66% to 67% for secondary education to 83–84% for lower than secondary and unknown education. Higher education, with an agreement rate of 72–73%, is placed in-between the latter figures. Table 2 suggests both a significant overstatement and understatement of education in the death records. Indeed, 27–28% of the death records with reported higher and secondary education correspond to lower self-reported education levels according to the census. And 5–6% of the deceased with secondary education as well as 16–17% of the deceased with an education lower than secondary or unknown are assigned in the death records to lower categories than according to the census information. These contradictory relationships suggest that some overstatement of education in the census records cannot be excluded.

Although the level of disagreement between the two sources of information (death and population census records) on marital status is generally lower

Table 2
Correspondence between data on education obtained from the death and census records, in percent

Education, census	Education, death records					
	Higher		Secondary		Lower than sec./unkn.	
	Males	Females	Males	Females	Males	Females
Higher	72.3	73.3	5.4	6.0	1.7	1.1
Secondary	20.5	21.2	66.4	67.3	15.4	14.8
Lower than sec./unkn.	7.2	5.5	28.2	26.7	82.9	84.1
Total	100	100	100	100	100	100

than that for education, the differences are still very substantial (Table 3). The highest accordance is found among married males and females, showing an agreement of 93% and 85%, respectively. Similar proportions are identified for those never married. The most significant difference is observed for the divorced and widowed. Among the male deaths, the agreement rate is 72% as to the divorced and 75% regarding the widowed. For females, the equivalent percentages are 62% and 90%, respectively (Table 3). The significant misreporting of the divorced status has also been observed in other countries (Weaver, 2000). This result may be partly affected by changes in marital status between the times of the census and death events due to the high divorce rate in Lithuania, with almost 60 divorces per 100 marriages in 2001–2004. However, we found that restricting our calculations to a shorter follow-up period (6 months) yields very similar results. So, we assume that changes in marital status during the period of follow-up have a relatively small effect on the disagreement rates.

Appendix B presents the statistical relationships between misreporting and selected socio-demographic characteristics, including age, cause of death, education, marital status, ethnicity, and the place of residence (urban/rural)). The outcomes of the Poisson regressions show that age is a significant predictor of reporting bias for education and, to some extent, for the female marital status. As for education, this relationship is more pronounced for females, with a risk of misreporting education at ages above 60 about 1.4–2.1 times higher than in the reference category (age group 30–39). With males, the age effects on misreporting are much less pronounced: only slight increases at ages 50–59 and 80+ are visible (the rate ratios of misreporting are 1.1 and 1.3, respectively).

Among the causes of death, deaths due to alcohol-related and ill-defined causes of death are associated with a 1.3–1.4 times higher risk of misreporting the marital status compared to deaths due to diseases of the cardiovascular system (reference category). The same causes of death also predict disagreement in information on education among males. Meanwhile, deaths from cancers are associated with a lower risk of misreporting the socio-demographic status (Appendix B).

The effects of education on misreporting the marital status are relatively small. In comparison to higher education (the reference category), an education lower than secondary is associated with about a 1.1 times higher risk of misreporting the male marital status.

As expected, the marital status is a strong predictor of reporting bias in the case of education. Those who reported themselves as never married in the census experience the highest risk of misreporting education in the death records (the misreporting rate ratios are 1.4 among males, and 1.3 among females) (Appendix B). An unknown marital status is a strong predictor of misreporting for females, whereas for males this relationship is not statistically significant. Smaller, but statistically significant effects are also identified for divorced and widowed males and females.

The deceased who identified themselves with an ethnicity other than Lithuanian (the reference category) in the census had a higher risk of misreporting their education and marital status in the death records. The effect is relatively small for misreporting education, but strikingly high for misreporting the female marital status: females belonging to the Russian and Other ethnicity groups had a risk of misreporting 1.2–1.5 times higher than those belonging to the Lithuanian ethnic group

Table 3
Correspondence between data on marital status obtained from the death and census records, in percent

Marital status, census	Marital status, death records							
	Married		Never married		Divorced		Widowed	
	Males	Females	Males	Females	Males	Females	Males	Females
Married	92.9	85.3	4.5	2.0	16.2	10.7	20.0	7.8
Never married	0.6	0.8	84.2	88.0	4.1	1.3	1.8	0.6
Divorced	4.5	3.2	7.7	2.8	72.5	61.8	3.0	1.8
Widowed	2.0	10.7	3.6	7.2	7.2	26.2	75.2	89.8
Total	100	100	100	100	100	100	100	100

(Appendix B). With the exception of the reporting bias of female education, the residents of rural areas had smaller probabilities of misreporting as to all of the three socio-demographic characteristics indicated in the death records.

There was virtually no correlation between the two types of misreporting (by education and marital status). The corresponding Pearson correlation coefficient was only 0.001. Differences in misreporting between marital status and education in respect to age and ethnicity show that they could have substantially different causes.

Socio-demographic differentials in life expectancy and relative mortality as reflected by the census-linked and unlinked data

Substantial intra-individual differences between the census and the death records described in the previous section produce differences between the census-linked and the unlinked group-specific mortality estimates. Fig. 1 compares the census-linked and unlinked group-specific mortality rates across ages. Looking closer at the trajectories of census-linked and census-unlinked group-specific mortality by age, we observe very similar patterns (Fig. 1). Although we found that reporting bias for education increases with age, the differences between the census-linked and census-unlinked mortality rates for higher and lower than secondary/unknown education categories are more pronounced at younger adult ages. This can be explained by the fact that group-specific mortality rates at these ages are sensitive to adjustments due to small numbers of deaths. In addition, combining unknown and lowest educational categories acts stronger at older ages towards diminishing the reporting bias.

Table 4 presents the census-linked estimates of group-specific life expectancies. It provides an opportunity to compare them with corresponding life expectancies calculated from the unlinked data.

According to the census-linked data, life expectancy at age 30 is 45.3, 39.3 and 34.4 years for men with higher, secondary, and lower and unknown education, respectively. So, the gap between the highest and the lowest educational categories is 10.9 years. As to women, life expectancies for the same categories are: 52.4, 49.0, and 45.6 years, and the gap is 6.8 years. For married males, life expectancy at age 30 constitutes 41.5 years, whereas that for the never married, divorced, and widowed males are close to each other and nearly equal to 31 year. The life expectancy of married females is 50.4 years and varies among single women from about 45 years (never married) to about 47 years (divorced).

Even before drawing a comparison with the census-linked data, one can notice unexpected directions of differences among life expectancies calculated from the unlinked data. There is an unexpected advantage of females with secondary education over females with higher education in terms of life expectancy after age 50.

The most significant disagreements between the census-linked and the unlinked life expectancy estimates are observed for education categories. The difference in male life expectancy at age 30 between the highest and lowest categories varies between 11 years (the census-linked data) and almost 15 years (the unlinked data). Both the higher life expectancy in the highest educational category and the lower life expectancy in the lowest educational category contribute towards diminishing the educational gap in the census-linked data (Table 4). For females, the disagreements between the census-linked and unlinked life expectancies are even greater. According to the census-linked and unlinked data, the differences between the highest and the lowest educational categories for females are 7 and 12 years, respectively. The unlinked estimates of life expectancy are about 3 years higher for higher education and about 2 years lower for the lowest education as compared to the corresponding census-linked estimates.

For marital status, the linked-unlinked differences in life expectancy at age 30 between the best (married) and the worst (widowed for males and never married for females) categories are relatively close to each other (Table 4). According to the unlinked data, the gap between the best and the worst categories constitutes more than 12 years for males and more than 6 years for females. This gap is over-estimated by almost one and half year for both males and females respectively when compared to the census-linked life expectancies.

Restricting our calculations to the upper age limit of 70 and applying temporary life expectancy to the exact ages 30–70 results in a relatively small impact on agreement between the census-linked and unlinked estimates (Table 4). The improvement affecting the unlinked estimates is more visible for educational differentials. First, the exclusion of the oldest ages allowed solving the problem of unreasonably high life expectancy for the secondary education group among females (especially visible

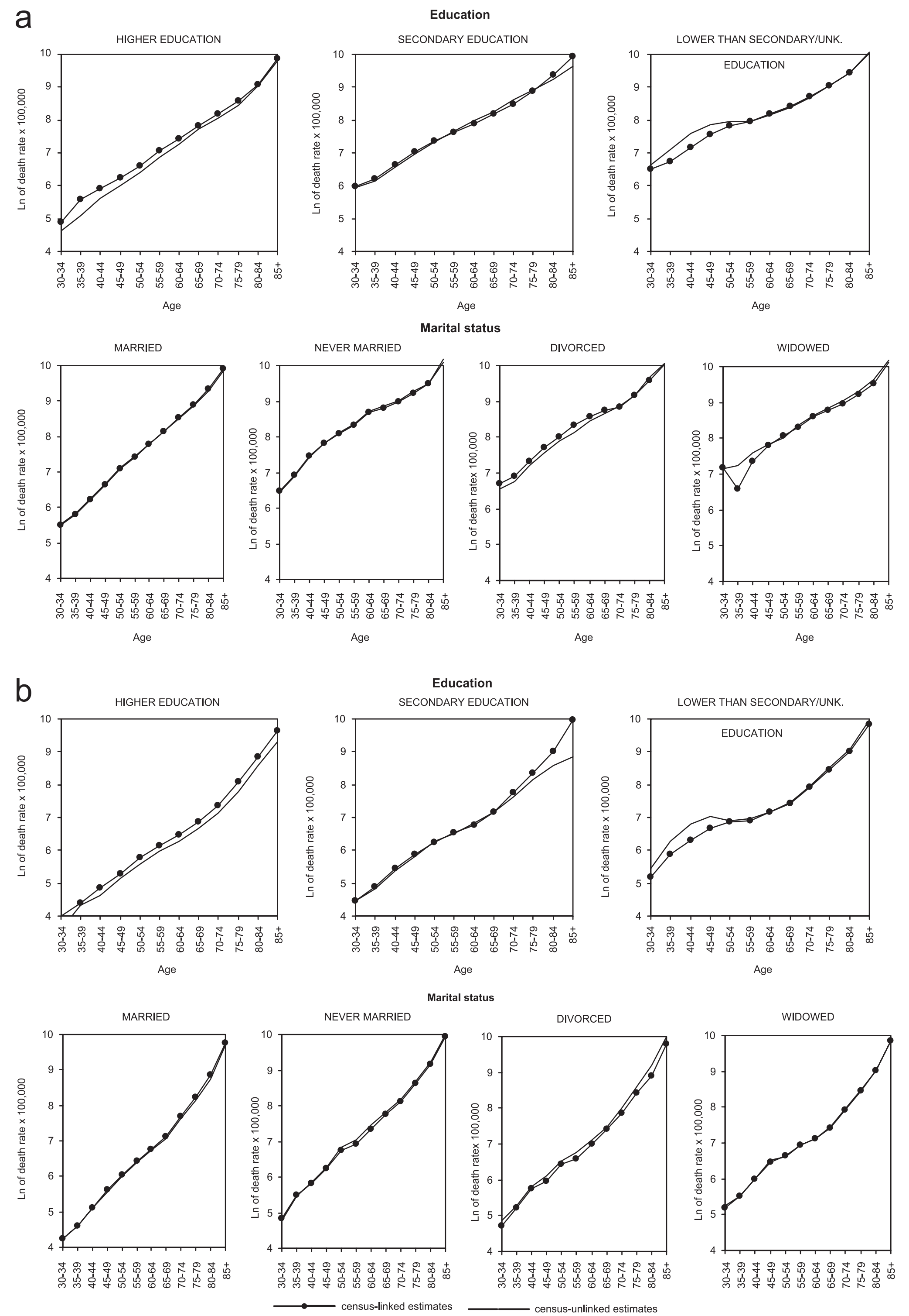

Fig. 1. Linked and unlinked age-specific mortality rates by education and marital status. (a) Males; (b) Females.

Table 4
Life expectancy at age 30 and temporary life expectancy between the exact ages 30 and 70, calculated according to the census-linked and the unlinked data

	Life expectancy at age 30				Temporary life expectancy (30–69)			
	Males		Females		Males		Females	
	Census-linked	Unlinked	Census-linked	Unlinked	Census-linked	Unlinked	Census-linked	Unlinked
Education								
Higher	45.28	47.02	52.38	55.27	36.30	37.02	38.49	38.74
Secondary	45.00–45.56	46.74–47.31	52.10–52.65	54.95–55.59	36.17–36.43	36.90–37.14	38.41–38.56	38.67–38.80
	39.27	39.42	49.00	54.54	33.26	33.36	37.67	37.71
	39.11–39.43	39.26–39.59	48.86–49.13	54.32–54.76	33.18–33.35	33.28–33.45	37.62–37.72	37.66–37.76
Lower than	34.42	32.34	45.58	43.60	30.28	28.59	35.71	34.61
secondary/unkn.	34.16–34.68	32.08–32.61	45.25–45.90	43.24–43.96	30.07–30.48	28.37–28.80	35.49–35.93	34.36–34.87
Highest–lowest	10.86[a]	14.68[a]	6.80[a]	11.67[a]	6.02[a]	8.43[a]	2.78[a]	4.12[a]
Marital status								
Married	41.49	41.51	50.44	51.09	34.82	34.70	38.04	38.09
	41.38–41.61	41.39–41.62	50.32–50.57	50.96–51.22	34.75–34.89	34.63–34.77	37.99–38.08	38.05–38.14
Never married	30.77	30.51	45.42	44.74	28.26	28.18	36.43	36.22
	30.45–31.10	30.20–30.82	45.07–45.76	44.40–45.09	28.03–28.49	27.95–28.40	36.25–36.62	36.04–36.41
Divorced	31.38	32.81	47.40	46.59	28.58	29.63	36.98	36.91
	31.08–31.67	32.51–33.11	47.11–47.69	46.32–46.87	28.36–28.80	29.41–29.85	36.86–37.10	36.79–37.03
Widowed	30.51	29.15	46.47	46.32	27.95	26.93	36.35	36.29
	29.09–31.93	27.74–30.56	46.09–46.86	45.93–46.71	26.69–29.22	25.66–28.19	36.08–36.61	36.02–36.56
Highest–lowest	10.98[b]	12.35[b]	5.03[c]	6.35[c]	6.87[b]	7.77[b]	1.69[b]	1.87[c]

[a]Higher–lower than secondary/unkn.
[b]Married–widowed.
[c]Married–never married.

after the age of 50). Second, the disagreement between the absolute differentials produced by the census-linked and unlinked estimates becomes less pronounced compared to life expectancy at age 30. However, the age restriction does not seem to be very helpful in solving problems around the unlinked longevity indicators by marital status.

Table 5 presents the relative group-specific mortality differentials calculated according to the census-linked and unlinked data. We found that Poisson mortality rate ratios derived from the two sources show similar patterns of inequalities both for education and marital status, with the most significant disagreements for males and females belonging to the lowest educational category. For example, the census-linked mortality for males and females with the lowest education was 2.2 and 1.8 times higher than for the higher education category. The corresponding relative mortality ratios reached 2.5 and 2.6 times according to the unlinked mortality data. The disagreements between the census-linked and unlinked estimates are much smaller in the case of relative mortality rates by marital status (Table 5). Generally, our data suggest that, although the unlinked mortality rate ratios for the worst-off education and marital status categories are likely to be over-estimated, they reveal the same tendencies as the census-linked estimates.

Summary of findings

This study used a unique data set from Lithuania containing aggregated mortality data based on the individual linkage of deaths in 2001–2004 to the census of 2001. In addition, the data include information about the socio-demographic characteristics of the deceased, obtained from the death certificate independently from the census information. This allowed us to produce the first census-linked estimates of life expectancy and mortality rate ratios by socio-demographic group for a post-Soviet country and to perform the first detailed analysis of the differences between the census-linked and unlinked estimates of group-specific mortality for an Eastern European country.

Table 5
Poisson regression mortality rate ratios for ages 30+ and for ages 30–69 calculated from the census linked and unlinked mortality data

	Poisson regression mortality rate ratios, age 30+				Poisson regression mortality rate ratios, ages 30–69			
	Males		Females		Males		Females	
	Census-linked	Unlinked	Census-linked	Unlinked	Census-linked	Unlinked	Census-linked	Unlinked
Education								
Higher	1	1	1	1	1	1	1	1
Secondary	1.59***	1.85***	1.53***	1.52***	1.76***	2.13***	1.49***	1.76***
	1.54–1.64	1.79–1.91	1.48–1.59	1.46–1.59	1.70–1.83	2.04–2.22	1.41–1.57	1.66–1.86
Lower than	2.15***	2.50***	1.75***	2.62***	2.63***	3.12***	2.30***	2.89***
secondary/ unkn.	2.08–2.21	2.42–2.58	1.68–1.82	2.51–2.73	2.53–2.74	3.00–3.26	2.18–2.43	2.72–3.07
Marital status								
Married	1	1	1	1	1	1	1	1
Never	2.03***	2.07***	1.63***	1.83***	2.39***	2.36***	1.84***	2.03***
married	1.97–2.08	2.01–2.13	1.59–1.68	1.78–1.89	2.31–2.46	2.29–2.44	1.74–1.94	1.92–2.14
Divorced	1.90***	1.68***	1.32***	1.49***	2.19***	1.84***	1.43***	1.52***
	1.86–1.95	1.63–1.72	1.27–1.36	1.44–1.54	2.13–2.26	1.79–1.90	1.37–1.50	1.46–1.59
Widowed	1.62***	1.79***	1.43***	1.54***	2.24***	2.31***	1.55***	1.63***
	1.58–1.66	1.75–1.83	1.40–1.45	1.51–1.57	2.15–2.34	2.22–2.40	1.49–1.61	1.57–1.70

All models are adjusted for age; ***$p \leqslant 0.001$; **$p \leqslant 0.01$.

This study is not affected by under-estimation of the overall and group-specific mortality due to the exclusion of the unlinked deaths, which was achieved by redistributing all unlinked deaths.

A comparison of the death record and census information about the same people discloses significant misreporting of education and marital status. The reporting biases are characterized by irregular patterns, with both an over- and under-representation of the socio-demographic categories. The misreporting of education and marital status tends to increase with age. Higher misreporting for alcohol-related and ill-defined causes of death are markers of the specific population groups characterized by greater disagreement between the census and death records. Higher disagreement rates are also associated with the status of divorcee and non-Lithuanian ethnicity. Finally, an urban place of residence increases bias in reporting education and marital status in Lithuania. The misreporting of education is unrelated to reporting bias in marital status.

The notable disagreement between the death record and census record information on the education of the same individuals can be attributed to several factors. First, authorized informants can provide imprecise information about the education of the deceased. It is even possible that education is imprecise in both data sources (Evans & McDonald, 1994). The second factor concerns transformations of the educational systems in Lithuania throughout the 20th century. A significant share of the people born before 1920 received educational degrees according to the Lithuanian and Polish rules (in the 1920s–1930s). The middle-aged generation of the Lithuanian population got their educational degrees following the soviet system practices (1940–1989). Finally, the youngest adults have experienced the transition from the soviet to the western type educational system in the 1990s. These changes cause difficulties in classifying educational degrees within the existing system. One can think, for example, that younger relatives trained within the soviet educational system face problems in reporting educational level of older people having degrees from the pre-war school. The third factor responsible for the discrepancy between the census and death records is related to the differences of educational categories. It seems that using eleven educational categories including some terms relevant to the prior systems reduces the probability of misreporting in the census records. However, objective information from educational registries would be a better option than self-reported census data.

The intra-individual biases in information about deceased result in significant differences between the census-linked and unlinked age-specific mortality rates. The most striking effect is shown for education, suggesting both a notable over-estimation of mortality for the lowest education below age 50 and an under-estimation of mortality for secondary education at the oldest age groups. The census-linked mortality rates for males and females with higher education are higher in all age groups compared to those derived from the death records. The two series of age-specific mortality estimates by marital status show a better agreement, suggesting smaller effects of misreporting.

The numerator-denominator bias affects both the absolute and relative measures of mortality inequalities. An exclusion of old ages improved the quality of unlinked estimates for females, but not for males. The study revealed that using death record based information on education and marital status leads to an over-estimation of life expectancy and relative mortality differentials.

A lack of census-linked data for 1989 makes it impossible to draw a definite conclusion about the reliability of the striking widening of socio-economic mortality differentials in post-Soviet countries over the 1990s reported in prior studies, including our own earlier works. According to our estimates, based on the census-unlinked data, the difference between the highest and the lowest educational categories grew from 9 and 5 years in 1989 to 15 and 12 years for males and females, respectively. There is no doubt that the widening was real and substantial, but its magnitude may be overestimated in the unlinked data. Indeed, the numerator–denominator bias was already notable in 1989 from the same (as in 2001–2004) implausible advantage of secondary education over higher education for females. The absence of census-linked data does not allow estimating directly how large the numerator-denominator bias was in 1989. As there were no major changes in the formal rules of statistical data collection during the 1990s, one can expect that the magnitude of the bias was the same in 1989 as it was in 2001.

The results of this study confirm the importance of educational and marital status mortality differentials in Lithuania. There are enormous mortality disadvantages in lower educated and non-married population groups which should be considered as a major obstacle for further health progress. Differential mortality trends should be monitored and addressed by adequate social and public health policies as a matter of urgency. Importantly, our study calls for producing better data for the assessment of differential mortality in Eastern Europe. In this connection, the fact that the population registers were launched in 1992 not only in Lithuania but also in Estonia and Latvia provides promising opportunities for the future.

Acknowledgement

This study has been completed as a part of the Vanguard Project of the Max Planck Institute for Demographic Research in Rostock (Germany) headed by James W. Vaupel. We are very grateful for Olga Trofimova (Statistics Lithuania) for her work on the initial version of the census-linked database. We are also very thankful for Danguolė Svidlerienė (Head of the Demographic Statistics division, Statistics Lithuania) for the most competent consultations on the mortality data in Lithuania. We extend our thanks to Jacques Vallin and France Meslé (INED) for their important contributions at the preparatory stages of the project. Finally, we are also very grateful for Susann Backer and Sigrid Gellers-Barkmann for help with language editing.

Appendix A

For each individual $k = 1, ..., N$, enumerated in the census, let $x_{ij}^k = 1$ if variable i (marital status, education, and ethnicity) has value j in the census record, and $x_{ij}^k = 0$ otherwise. For unlinked deaths, x_{ij}^k is 0 for any pair (i, j).

Let $\hat{x}_{ij}^k$ be equal to x_{ij}^k for all records except unlinked deaths. To assign some values to $\hat{x}_{ij}^k$ for unlinked deaths, the following method is applied. First, the probabilities of being in a certain census-based category of education, marital status, and ethnicity conditioned on age, year, place of residence, cause of death, and of being in the death-record-based categories of the three socio-demographic variables are estimated by multinomial logistic regression. Thus, we estimated the probabilities of having each census-based category for the census-linked deaths:

$$P_{i,j,k} = P(x_{ij} = 1 | \text{age, year, place of birth, place of residence, cause of death, } x_{ij}^{\text{certificate}}),$$

where $x_{ij}^{\text{certificate}}$ is a variable, similar tox_{ij}, based on death certificate.

Second, we assigned the conditional probabilities of having census-based education, marital status, and ethnicity for each unlinked death. For each census-unlinked death k, the initial values $\hat{x}_{ij}^{k}$ are chosen randomly according to probabilities $P_{i,j,k}$, corresponding to socio-demographic categories.

The final assigned values $\hat{x}_{ij}^{k}$ are estimated as values producing the minimum values for the functional:

$$\sum_{i=1}^{3}\sum_{j=1}^{J_i}\left(\frac{\sum_{k=1}^{N}\hat{x}_{ij}^{k}}{\sum_{k=1}^{N}\sum_{j=1}^{J_i}\hat{x}_{ij}^{k}}-\frac{\sum_{k=1}^{N}x_{ij}^{k}}{\sum_{k=1}^{N}\sum_{j=1}^{J_i}x_{ij}^{k}}\right)^2\rightarrow\min_{\hat{x}_{ij}},$$

where 3 is the number of socio-demographic variables, J_i is the total number of the possible values (categories) of the socio-economic variable *i*. The minimization problem has to be solved for every age group and cause of death. In doing so, we applied the random search method (see Zhigljavsky (1991), for example). As a result of this estimation procedure, the census-based education, marital status, and ethnicity are assigned to every census-unlinked death with producing a minimum change in the distributions of deaths by socio-demographic categories.

Appendix B

Relationships between misreporting of education, marital status and socio-demographic factors. Outcomes of Poisson regressions (Table B1).

Table B1

	Misreporting of education		Misreporting of marital status	
	Males	Females	Males	Females
Age group				
30–39	1	1	1	1
40–49	0.98	1.01	1.04	1.06
	0.91–1.05	0.87–1.19	0.94–1.16	0.85–1.31
50–59	1.10*	1.29*	1.08	1.21
	1.03–1.18	1.11–1.49	0.98–1.20	0.99–1.48
60–69	1.03	1.42*	0.91	1.33*
	0.96–1.10	1.24–1.63	0.83–1.01	1.10–1.60
70–79	1.03	1.45*	0.82*	1.27*
	0.97–1.10	1.27–1.67	0.74–0.90	1.06–1.53
80+	1.27*	2.10*	0.93	0.90
	1.19–1.36	1.83–2.40	0.84–1.03	0.75–1.09
Cause of death				
Cardiovascular syst. dis.	1	1	1	1
Neoplasms	0.94*	0.88*	0.71*	0.96
	0.91–0.97	0.84–0.91	0.67–0.75	0.91–1.02
External causes	1.00	0.97	1.13*	1.06
	0.96–1.05	0.90–1.04	1.05–1.20	0.96–1.18
Alcohol-related deaths	1.14*	1.03	1.27*	1.34*
	1.06–1.23	0.92–1.15	1.14–1.41	1.16–1.55
Infectious and respiratory system diseases	1.03	0.99	1.10*	1.04
	0.98–1.09	0.92–1.06	1.02–1.19	0.92–1.18
Ill-defined causes	1.21*	1.12	1.44*	1.35*
	1.05–1.39	0.92–1.37	1.19–1.74	1.01–1.78
All other causes	1.01	0.97	0.99	1.07
	0.95–1.08	0.92–1.03	0.90–1.10	0.98–1.17
Education				
Higher			1	1
Secondary			1.06	1.04
			0.98–1.16	0.94–1.15
Lower than secondary			1.11*	0.99
			1.02–1.21	0.90–1.09

Table B1 (*continued*)

	Misreporting of education		Misreporting of marital status	
	Males	Females	Males	Females
Unknown			1.17*	0.85*
			1.05–1.30	0.76–0.95
Marital status				
Married	1	1		
Never married	1.38*	1.33*		
	1.32–1.45	1.27–1.40		
Divorced	1.21*	1.11*		
	1.16–1.27	1.05–1.19		
Widowed	1.14*	1.14*		
	1.10–1.18	1.10–1.18		
Unknown	1.56	1.92*		
	0.94–2.59	1.37–2.69		
Ethnicity				
Lithuanian	1	1	1	1
Russian	1.17*	1.16*	1.24*	1.47*
	1.11–1.22	1.11–1.22	1.15–1.33	1.37–1.57
Polish	1.11*	1.11*	1.01	1.16*
	1.05–1.16	1.06–1.16	0.94–1.09	1.08–1.25
Other	1.18*	1.11*	1.16*	1.40*
	1.10–1.26	1.04–1.19	1.04–1.30	1.26–1.55
Unknown	2.02*	2.21*	0.89	1.11
	1.61–2.54	1.75–2.79	0.53–1.47	0.65–1.87
Place of residence				
Urban	1	1	1	1
Rural	0.89*	1.03*	0.95	0.77*
	0.87–0.91	1.01–1.06	0.92–1.00	0.73–0.80

All models are adjusted for age; $^{*}p \leqslant 0.05$.

References

Andreev, E., & Dobrovolskaya, V. (1993). Sotsio-kulturniye razlichiya v smertnosti v Rossii [The socio-cultural differences in mortality in Russia]. *Zdravoohraneniye Rossiiskoi Federatsii*, *12*, 18–21.

Blazek, J., & Dzúrová, D. (2000). The decline of mortality in the Czech Republic during the transition: A counterfactual case study. In G. A. Cornia, & R. Paniccià (Eds.), *The mortality crisis in transitional economies* (pp. 303–327). Oxford: Oxford University Press.

Chiang, Ch. L. (1984). *The life table and its applications*. Malabar, FL: Robert E. Krieger Publishing Company.

Doblhammer, G., Rau, R., & Kytir, J. (2005). Trends in educational and occupational differentials in all-cause mortality in Austria between 1981/82 and 1991/92. *Wien Klin Wochenschr*, *117*(13&14), 468–479.

Evans, J., & McDonald, D. (1994). *1991 census data quality: Education. Census Working Paper 94-2*. Canberra: Australian Bureau of Statistics.

Goldblatt, P. (1989). Mortality by social class, 1971–85. *Population Trends*, *56*, 6–15.

Huisman, M., Kunst, A. E., Bopp, M., Borgan, J.-K., Borrell, C., Costa, G., et al. (2005). Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. *Lancet*, *365*, 493–500.

Kalediene, R., & Petrauskiene, J. (2000). Inequalities in life expectancy in Lithuania by level of education. *Scandinavian Journal of Public Health*, *28*(1), 4–9.

Kalediene, R., & Petrauskiene, J. (2005). Inequalities in mortality by education and socio-economic transition in Lithuania: Equal opportunities? *Public Health*, *119*, 808–815.

Klinger, A. (2001). Halandósági különbségek maguarországon iskolai végzttség szerint [Mortality differences by educational level in Hungary]. *Demografia*, *44*, 227–258.

Kohler, I. (2001). Adult and old-age mortality dynamics in Bulgaria and Russia (Dissertation). Odense: Faculty of Social Sciences, University of Southern Denmark.

Kunst, A.E., Bos, V., Santana, P., Valkonen, T., Mackenbach, J.P., Andersen, O., et al. (2004). Monitoring of trends in socioeconomic inequalities in mortality: Experiences from a European project. Demographic Research, Special Collection 2 (Determinants of Diverging Trends in Mortality), pp. 229–254.

Leinsalu, M., Vågerö, D., & Kunst, A. E. (2003). Estonia 1989–2000: enormous increase in mortality differences by education. *International Journal of Epidemiology*, *32*, 1081–1087.

Leinsalu, M., Vågerö, D., & Kunst, A. E. (2004). Increasing ethnic differences in mortality in Estonia after the collapse of the Soviet Union. *Journal of Epidemiology and Community Health*, *58*(7), 583–589.

Marmot, M. G., & McDowall, M. E. (1986). Mortality decline and widening social inequalities. *Lancet*, *2*, 274–276.

Mathers, C. D., Fat, D. M., Inoue, M., Rao, Ch., & Lopez, A. D. (2005). Counting the dead and what they died from: An assessment of the global status of cause of death data. *Bulletin of the World Health Organization*, *83*, 171–177.

Petrauskienė, J., Bierontas, D., Kalėdienė, R., & Zaborskis, A. (1996). *Lietuvos gyventojų mirtingumo medicininiai socialiniai aspektai [Medical and social aspects of mortality of the Lithuanian population]*. Kaunas: Kaunas Medical Academy.

Plavinski, S. l., Plavinskaya, S. I., & Klimov, A. N. (2003). Social factors and increase in mortality in Russia in the 1990s: prospective cohort study. *British Medical Journal*, *326*, 1240–1242.

Rosamond, W. D., Tyroler, H. A., Chambless, L. E., Folsom, A. R., Cooper, L., & Conwill, D. (1997). Educational Achievement Recorded on Certificates of Death Compared with Self-Report. *Epidemiology*, *8*(2), 202–204.

Shai, D., & Rosenwaike, I. (1989). Errors in reporting education on the death certificate: some findings for older male decedents from New York State and Utah. *American Journal of Epidemiology*, *130*(1), 188–192.

Shkolnikov, V. M., Deev, A. D., Kravdal, Ø., & Valkonen, T. (2004). Educational differentials in male mortality in Russia and Northern Europe. A comparison of an epidemiological cohort from Moscow and St. Petersburg with the male populations of Helsinki and Oslo. *Demographic Research*, *10*, 1–26.

Shkolnikov, V., Leon, D., Adamets, S., Andreev, E., & Deev, A. (1999). Educational level and adult mortality in Russia: An analysis of routine data 1979 to 1994. *Social Science & Medicine*, *47*(3), 357–369.

Sorlie, P. D., & Johnson, N. J. (1996). Validity of education information on the death certificate. *Epidemiology*, *7*(4), 437–439.

Stalioraityte, E., Pangonyte, D., & Kazlauskaite, D. (2005). Reliability of cause-specific mortality rate statistics: Case of Lithuania. *Public Health*, *119*, 799–807.

Stankuniene, V., Mikulioniene, S., & Baublyte, M. (2002). Gyventojų skaièius ir sudėtis. *Gyventojų senėjimas* [Population size and its composition. Population ageing]. In V. Stankuniene (Ed.), *Lietuvos gyventojai* (pp. 7–24). Vilnius, Lithuania: Demographic Research Centre (LIPS).

Valkonen, T. (1993). Problems in the measurement and international comparisons of socio-economic differences in mortality. *Social Science & Medicine*, *36*(4), 409–418.

Vallin, J. (1979). Socioeconomic determinants of mortality in industrialized countries. *Readings in Population Research Methodology*, *2*, 957–971.

Weaver, D. A. (2000). The accuracy of survey-reported marital status. Evidence from survey records matched to Social Security records. *Demography*, *37*(3), 395–399.

Zhigljavsky, A. A. (1991). *Theory of global random search*. Dordrecht/Boston/London: Kluwer Academic Publishers.